R E P O R T  R E S U M E S

ED 011 692 EA 000 276

MERIT RATING FOR SALARY INCREASES AND PROMOTIONS.

BY- BOLIN, JOHN G. MUIR, JOHN W.
GEORGIA UNIV., ATHENS, INST. OF HIGHER EDUCATION

PUB DATE 66

EDRS PRICE MF-$0.09 HC-$1.68 42P.

DESCRIPTORS- *HIGHER EDUCATION, TEACHER EVALUATION, *FACULTY EVALUATION, *MERIT PROGRAMS, *FACULTY PROMOTION, *RATING SCALES, EVALUATION TECHNIQUES, ADMINISTRATIVE PERSONNEL, PEER GROUPS, STUDENT OPINION, SELF EVALUATION, BEHAVIOR RATING SCALES, EFFECTIVE TEACHING, ATHENS

DEMANDS FOR QUALITY EDUCATION HAVE EMPHASIZED THE NEED FOR OBJECTIVE MERIT RATING OF TEACHERS FOR PROMOTION AND SALARY INCREASES. THE STUDY REVIEWS MERIT-RATING LITERATURE, ANALYZES RATING SCALES AND PROCEDURES, AND MAKES RECOMMENDATIONS FOR DEVELOPING AND IMPLEMENTING MERIT PROGRAMS. THE WRITERS CONCLUDE THAT A COMBINATION OF SUPERORDINATE, PEER, STUDENT, AND SELF-EVALUATION BEST PROVIDES A BASIS FOR DECISIONS ON MERIT INCREASES AND PROMOTIONS. CONTENTS OF 21 RATING SCALES (SEVEN SUPERORDINATE, FOUR PEER, FIVE SELF, AND FIVE STUDENT) REVEALED THAT THE IMPORTANT FACTORS INVOLVED IN FACULTY RATING ARE, IN DESCENDING ORDER, (1) CLASSROOM TEACHING, (2) PERSONAL ATTRIBUTES, (3) PROFESSIONAL GROWTH ACTIVITIES, (4) STUDENT-FACULTY RELATIONS, (5) COMMUNITY SERVICE, (6) SERVICE TO THE INSTITUTION, (7) LENGTH OF SERVICE, (8) RESEARCH AND PUBLICATION, AND (9) COMPETING OFFERS. THIS ORDERING SHOWS A HIGH CORRELATION WITH PREVIOUS STUDIES. THERE IS NO STANDARD PATTERN OF MERIT RATING. HOWEVER, THE AGREEMENT ON FACTORS IMPORTANT IN JUDGING TEACHER EFFECTIVENESS CAN SERVE AS A GUIDELINE IN DEVELOPING A MERIT PROGRAM. SEVERAL RATINGS FROM DIFFERENT SOURCES SHOULD BE USED, AND BOTH DESCRIPTIVE PHRASES AND NUMERICAL RATING ARE USEFUL. THIS PUBLICATION IS ALSO AVAILABLE FROM THE INSTITUTE OF HIGHER EDUCATION, UNIVERSITY OF GEORGIA, ATHENS, GEORGIA, FOR $.50. (HM)

ED011692

# MERIT RATING FOR SALARY INCREASES AND PROMOTIONS

John G. Bolin and John W. Muir

INSTITUTE OF HIGHER EDUCATION
University of Georgia
Athens, Georgia

1966

EA000276

# STAFF

## INSTITUTE OF HIGHER EDUCATION

University of Georgia
Athens, Georgia

| | |
|---|---|
| GALEN N. DREWRY | Director |
| CAMERON FINCHER | Associate Director |
| JOHN G. BOLIN | Assistant Professor |
| TED L. HAMMOCK | Instructor. Supervisor, FAIT Project |
| JOHN J. SULLIVAN | Instructor. Supervisor, FAIT Project |
| WILLIAM W. PARKER | Coordinator, Administrative EDP |
| DONALD E. DOWDLE | Systems Analyst/Programmer |
| DANNY W. SHEALY | Systems Analyst/Programmer |
| JOHN MUIR | Research Assistant |
| BOB CRAWFORD | Research Assistant |
| BYRON JAY | Research Assistant |
| NICHOLAS ALIOTTI | Research Assistant |

Secretaries

| | |
|---|---|
| Mary Snyder | Kathleen Coy |
| Barbara Brown | Margaret Snyder |

# MERIT RATING
# FOR
# SALARY INCREASES AND PROMOTIONS

by
John G. Bolin
John W. Muir

INSTITUTE OF HIGHER EDUCATION
University of Georgia
Athens, Georgia

1966

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Copyright © 1966
by
Institute of Higher Education
University of Georgia

"PERMISSION TO REPRODUCE THIS
COPYRIGHTED MATERIAL HAS BEEN GRANTED
BY Galen N. Drewry

TO ERIC AND ORGANIZATIONS OPERATING
UNDER AGREEMENTS WITH THE U.S. OFFICE OF
EDUCATION. FURTHER REPRODUCTION OUTSIDE
THE ERIC SYSTEM REQUIRES PERMISSION OF
THE COPYRIGHT OWNER."

TABLE OF CONTENTS

Background.....1

Developing The Merit Plan.....4

The Merit Concept.....4

Identifying the Criteria.....6

Constructing the Instrument.....10

The Evaluator.....12

Merit Programs--Success or Failure.....17

A Comparison of the Contents of Rating Scales....19

Implications.....24

Summary and Conclusions.....26

Bibliography.....30

Appendix.....33

# FOREWORD

The study on merit pay and merit promotion programs was proposed and partially financed by the Georgia Association of Junior Colleges. Appreciation is expressed to that Association for its interest in and support of the study.

This study provides a review of current literature regarding many of the issues surrounding the merit rating concept. It also makes a brief comparative analysis of twenty-one rating scales for the existence of differences and similarities. While the study is not intended to be an exhaustive research report, nor to point to any final answers to the merit issue, it has attempted to provide helpful suggestions and guidelines for developing a sound program based on merit. The final responsibility for any merit rating plan must, nevertheless, lie within the administration and faculty of each institution.

JGB
JWM

# MERIT RATING
# FOR
# SALARY INCREASES AND PROMOTIONS

## Background

Merit pay and merit promotions have long been controversial issues in the education profession. The growing shortage of professors in higher education has caused attention to be focused on merit programs as a means to recruit top quality faculties.

Particularly within the past several years, the concept of merit ratings for pay and promotion has become one of the most widely discussed problems circulating among institutions of higher education. It seems to have grown out of the demands of the public for quality education in the colleges and universities of today. These demands have been echoed loudly from coast to coast. Greater numbers of students are demanding more and better education than ever before.

Education, as well as knowledge, is growing both in breadth and depth. The development of new techniques and new ideas in all areas of learning requires well prepared, well qualified, and up-to-date instructors. No longer can the professor stimulate his students using as his only reference lecture notes made three or four years ago. Modern institutions of higher learning have no room for the lackadaisical pedagogue of yesteryear. Quality and performance are the essence of today's college programs.

Encouraging quality in the instructional program is one of the most complex and perplexing problems facing the college administrators and their faculties. It is one to which there is no easy solution. One means often used by administrators to encourage quality is to reward excellence or to recognize superior effort or achievement through advancement.

Up to this point, however, the recognition of superior performance, on the part of faculty members, has been accomplished more by guesswork than by any other means. The decisions and judgments made by administrators with regard to pay increases and promotions for faculty members have been based, to a large extent, on hearsay or gossip. To be successful, however, a merit program cannot operate as a "hit-and-miss" proposition. To be effective, there must be some consistency, some design or set of guidelines to assist the administrator or the evaluator in identifying outstanding individuals, and those faculty members whose performance merits extra consideration. Measuring the work of individuals against some established criteria would tend to reduce, or minimize the guesswork in determining who should receive advancements in pay and/or position.

Salary increases and promotions based upon merit ratings are not new in our competitive society. Merit rating has taken place in business and industry for many years. Incentive raises and incentive promotions are well known in the world of business. In industry workers are often rated by production standards, and then paid accordingly. Those workers who produce the most units and the best quality are usually rewarded for their efforts through advances in salary and position. This practice also continues on up the line; foremen and supervisors, too, are rated by their superiors according to their production output as compared with the performance of other foremen and supervisors, and salary increases and promotions are awarded in line

with these ratings. Even presidents and vice-presidents of large manufacturing companies are evaluated. They are rated both by the public and by the stockholders. They are rated by the public in terms of the demands for the product. By their stockholders, they are rated on the basis of the profit or loss which their company shows annually. So, in effect, they receive two ratings annually, both having a bearing on their future.

Rating of professional people is accomplished on a somewhat less formal basis. Because of the more intangible nature of the professional world, these ratings consist more of opinions rather than of facts. A physician, for example, may be rated with regard to how he treats his patients, not only physically but also socially. Some doctors are rated in terms of the lack of success, unfortunately for some of us, which they have had in curing illnesses. Dentists may be ranked in terms of how comfortable a patient is made to feel and not necessarily on how well he cares for the patient's teeth. Lawyers, on the other hand, are assessed in terms of success they have in settling disputes, both in and out of the courtroom. These ratings while perhaps less formal than those in the industrial world are sometimes more effective than those which allow the use of a checklist.

Another example is found in the military services. The life of the career serviceman, particularly in the officer ranks, literally hangs in the balance of merit ratings as was illustrated quite clearly during the "Reduction in Force" in 1957-58. In the military a serviceman's performance is rated annually by his superiors and reviewed periodically, and he is promoted and given assignments in accordance with the results of these rating sheets.

In every facet of society comparisons and evaluations are continuously being made. Some are made objectively, some subjectively, and others by a mixture of opinions and facts; nevertheless, they are made. The ability to discriminate virtually compels us to make comparisons and to rank people in order of importance and value. All, of course, may not agree with the precise order or rank of importance, but each person has some system or method of rating others.

The following suggestions, therefore, attempt to provide a framework for rating so that it may be accomplished in some organized or orderly fashion. It is firmly believed that systematic rating will ultimately improve the accuracy of judgments and decisions presently being made on the basis of guesswork.

## Developing The Merit Plan

### The Merit Concept

The philosophy behind merit pay and merit promotion plans is not usually a topic for argument among educators. Just as professors have recognized individual differences in students, there exist individual differences among faculty members. There are differences of motivation, initiative, interest, willingness, loyalty, and persistence. The variations in these factors, and others influence the individual's productivity. The quality of work produced as well as the quantity will vary according to the interaction of these elements within the individual. Yet, even with this awareness of individual differences, there is strong opposition to implementing merit programs in institutions of higher education.

Merit ratings are opposed frequently because of the assumed ill-effects they have on instruction. It is commonly believed that a rating checklist used by a department head or dean will force teachers to conform to policies and methods devised by the administrators themselves. This argument has been expressed best by Wiles (1955: 294) who stated that "rating brings with it a reduction of the freedom of the teacher and the class to follow the learning procedures that seem most profitable to them."

Yet there are few, if any, who would permit an incompetent faculty member to have complete freedom to teach. Is it logical that the faculty should be granted the freedom to teach whatever they wish however they please? Should there not be certain instructional guidelines in force in every institution of higher learning? These questions do not imply that each faculty member must sit or stand in a specified position; nor do they imply that lesson plans must be outlined and followed to the last minute detail. Nevertheless, college teachers must maintain some dignity in the classroom, attempt some organization of instruction, and direct instruction toward some objectives which have been identified.

College administrators are charged with the responsibility of selecting competent faculty members and insuring that instruction in their institutions meets some prescribed standards. It is not logical, therefore, to refuse to recognize differences in teaching effectiveness, to rate the effectiveness of a professor on the basis of unfounded rumor or hearsay. The administrator must have some consistent basis for making his judgments.

## Identifying the Criteria

The crux of the problem in merit plans is the basis for determining merit. The difficulties in identifying a common and concise set of standards on which to base a sound merit program, have been frustrating to numerous educators. Opponents to merit pay are quick to point out that no valid crieria as yet have been defined. Even the "experts," they say, do not agree upon determinants of teaching effectiveness. (Brueckner 1955: 343) Mitzel (1960: 1481) noted in 1957 that "no standards exist which are commonly agreed upon as the criteria of teacher effectiveness." Yet, "any attempt at evaluation of teacher effectiveness must first of all deal with the problem of an adequate criterion." (Yamamoto, 1963: 31)

Criteria used to evaluate teaching effectiveness have been diverse, and descriptions of traits or characteristics are multitudinous. In his review of rating scales, Barr (1948: 213-314) identified 200 different traits which were included in 209 different rating scales. Many of these, however, covered the same general area, the only differences being in the wording of the trait or characteristic. For instance, Barr's list included "personal appearance," "appearance," "dress," "general appearance," "personal characteristics," all of which could be tabulated under the general heading of personal appearance. Still another example includes the scope of preparation. Overlapping items such as "lesson planning," "preparation," "daily preparation," "organization of subject matter," "knowledge of subject matter," "preparation of daily work," and "preparation of work," were included separately.

This diversity in trait descriptions does not mean that there is complete lack of agreement in what is to be rated. In fact the 200 entries listed by Barr could be easily reduced to perhaps ten or

fifteen areas depending on what classification one would care to make. For example, one could establish factorial categories such as "personal habits or characteristics," "teaching techniques," "academic or professional growth." By reducing the duplication of items one would be able to note considerable similarity among institutions of the same type. It seems, then, that the only logical reason for the variety of ways an item is expressed on the different rating scales is that the explicit terminology used in each item was the most meaningful for the situation in which those particular individuals were involved.

Institutions of higher education throughout the nation strive toward a multitude of diverse objectives. The divergence in settings and goals is the major reason why no single overall plan for evaluating college faculty members exists. The diversity of roles assumed by different institutions has a distinct bearing on what is considered important and how much weight should be placed on the factors included in the evaluation. Many colleges and universities will place considerably more emphasis upon lower division instruction, while others prefer to concentrate their efforts in research at the graduate level. Two reports of surveys of various types of institutions of higher education* conducted for the American Council on Education (Gustad, 1961, and Astin and Lee, 1966: 347-375), however, indicated that in rank of importance, classroom teaching, personal characteristics, and student advising were almost universally the most important factors to be evaluated. Several other criteria such as committee

---

* Classification of types included: Liberal Arts Colleges, Private Universities, State Universities, State Colleges, Teachers Colleges, Junior Colleges, and Professional and Technical Colleges.

work, research, and publication, however, varied in rank among the different types of institutions. Research, for instance, was weighted heavily in the private and state universities, but not among the other types of institutions. Publication was considered less important in the junior colleges, the teacher colleges, and the state colleges than in the universities and technical colleges.

Yet, while these and some other differences may exist among various types of institutions of higher education, agreement within similar types seems to be prevalent. Specifically regarding junior colleges, Gustad found that with regard to frequency of use, teaching was indisputably the most important single factor in faculty evaluation; (See Table 1) publication and research, on the other hand, were decidedly less important. These results were substantiated five years later by Astin and Lee. They also found teaching to be the most frequently used factor in evaluation of faculty members for promotion, salary increases, and tenure. Teaching was followed by personal attributes, student advising, and committee work in rank order of importance. In the overall comparison of factors used for evaluation there is a strong agreement among the factors included in the two studies. A rho of more than .90 computed for the two lists implies that over the past several years a consistent pattern has developed among junior colleges regarding the importance assigned to the various factors used to evaluate faculty effectiveness. In support of the findings reported by Gustad, and those of Astin and Lee, Graybeal (1966: 48-49) has indicated that over 70 percent of the junior college faculty members included in his study believed that teaching ability was considered more important than was publishing.

These studies indicate that although no set of standards exists which all institutions can apply uniformly, there is considerable agreement among similar institutions as to the relative

TABLE 1
FACTORS FOR FACULTY EVALUATION

| | Factors | Rank | |
|---|---|---|---|
| | | Gustad | Astin & Lee |
| 1. | Classroom Teaching | 1 | 1 |
| 2. | Other*, Personal Attributes | 2 | 2 |
| 3. | Student Advising | 3 | 4 |
| 4. | Committee Work | 4 | 5 |
| 5. | Length of Service in Rank | 5 | 3 |
| 6. | Professional Society Activity | 6 | 6 |
| 7. | Public Service | 7 | 7 |
| 8. | Publication | 8 | 11.5 |
| 9. | Research Activity | 9 | 9 |
| 10. | Competing Offers | 10 | 10 |
| 11. | Supervision of Honors | 11 | 8 |
| 12. | Consultation | 12 | 11.5 |
| | Rho = .90 | | |

* This category, it was explained in the Gustad report, should be included with "personal attributes" since the responses tended to overlap considerably.

importance of certain general factors used in faculty ratings. Perhaps even more important, these studies suggest that most institutions tend to rely upon the internal development of a rating plan. That is, factors which are to be used for faculty evaluation are most likely to be derived by some person or committee within the institution. Those on the local scene naturally understand the particular function or role of their college better than outsiders. This does not mean that outside advice and counsel should not be obtained; it simply means that ultimately the best plan for evaluating the members of a given faculty must come from the members themselves, using research and literature only as guidelines to assist them in their efforts.

## Constructing the Instrument

Deciding what factors are important in teaching effectiveness is but one step in developing the merit plan. Regardless of how refined or explicit the criteria may be, the construction of a good rating instrument may prove to be a major problem.

Rating scales seem to be the most common formal device used to assess effectiveness and thereby determine merit. On the surface they are the easiest to apply and to interpret. The rater seemingly has a simple task because he is required merely to assign a numerical or letter value to the criteria included on the rating form. Because of the ordinal descriptions these rating scales may be analyzed with ease and pose little difficulty for the reviewer in comprehending the rank value of each item.

The rating scale, however, is not without its pitfalls. An obstacle for the rater may become apparent when he must make a judgment as to the numerical value of an item based on his observations of the instructor, the teaching environment, his students, and/or from a review of records pertaining to the individual and his work. In some cases this does not really get to the heart of the matter. That is, a numerical value may not really present the complete picture of how well or how poorly an individual performs. Perhaps the old axiomatic expression that "the whole is greater than the sum of its parts" has meaning in this case.

Other merit plans, on the other hand, have relied solely upon narrative evaluations--word descriptions of performance, such as: "Mr. Jones displays a professional interest in teaching;" or "Mr. Brown encourages class participation." While these statements may not be found on an actual evaluation form, they illustrate the potential problem

for the reviewer. What is meant precisely by "professional interest" or "insures class participation?" Did the rater feel that Mr. Jones displayed a greater interest in teaching than did Mr. Brown; or did he feel that Mr. Brown's class had more or less participation than Mr. Jones' class? Moreover, does the rater mean that participation is synonymous with interest? These questions may not be answered in the narrative description.

It seems then, that both types of rating techniques have advantages and disadvantages. Because neither can do the best job, the two techniques frequently are used in combination. While there is no question that a combined approach will mean more work on the part of the evaluator, a more complete and accurate rating may be obtained.

One weakness common to nearly all rating scales lies in the instructions which are supposed to facilitate interpretation of items. Great care should be taken to insure that good explanatory statements are included with each item. Frequently these statements are too parsimonious, failing to provide an adequate description of the behavior the rater must judge. In other instances the definitive statements may be too broad or general to provide the rater with a sound basis for making a judgment. Carefully thought-out and precisely worded definitions therefore are extremely important to the effectiveness of an evaluation.

Ultimately it must be remembered that there is no completely objective method of evaluating human behavior. The inconsistency of environment and the never-ceasing changes within individuals preclude a truly objective analysis or rating of others. Progress can be made, nevertheless, through the continued effort to refine and improve the techniques used in making judgments and in rating faculty competence.

## The Evaluator

If agreement has been reached on criteria which are in line with the functions of the institution, and if agreement has been reached on the instrument to be used for the evaluation, a third and possibly most difficult task still remains. Who should make the evaluation? In other words who are "...competent judges or raters..."? (Engleman, 1957: 138)

Superordinate Ratings. Probably the most commonly used evaluation scheme is that of superordinate rating. In this approach the department head or dean generally has the responsibility of evaluating faculty members. Ratings of this type generally require the rater to observe the performance of the individual in the classroom and to review the individual's personnel files or records. As a final step in the evaluation, the rater may conduct a conference in which the results of the evaluation will be discussed.

This procedure for evaluation, while somewhat inconvenient to both faculty and college administrators, is usually fairly effective and has a number of advantages over other procedures. Not only does it allow the rater at least two sources of information on which he can base his evaluation, but more important it provides an opportunity for a face to face discussion where differences may be examined and where weaknesses may be identified and corrections suggested. It requires little time on the part of the individual faculty member. Furthermore the form on which the rating is made is usually completed in objective terms.

The chief argument against this approach is that the accuracy of the rating is dubious. That is, the rating usually is based upon only one or, at best, few observations. Since effectiveness will

fluctuate from day to day, it is maintained that it is not a fair evaluation of total effectiveness, that it is an inadequate sample on which to base a judgment. Hence, although evaluation in this manner may tend to conserve time, it does not provide a sufficient amount of information to be used as the sole basis for determining merit for promotion or salary increases.

Peer Rating. A second approach to the problem of identifying competent judges, and one which appears to be more acceptable to those judged, is to use ratings by colleagues or peers. For the competent faculty member, peer ratings do not create quite so much threat as ratings by department heads or deans. Peer evaluation also has the advantage of having a rater who is familiar with the field or subject area of the person being rated. Peer ratings are effective, however, only when there is adequate communication among faculty members. Where communications are lacking, peer judgments become less reliable. Slocum, (1965) for instance, found that peers having two-way communications with other teachers were considered better teachers by the in-group than by those with whom they had no real contact.

Of course knowledge of the subject does not necessarily imply effectiveness in the classroom. It is not too uncommon for some faculty members to "talk a good game." A physics professor, for example, may be able to devise a new technique for reducing the drag factor to increase the lift factor on a space vehicle, but he may be the drag factor himself in the classroom. On the other hand, a history professor who is considered by his peers to have only an average knowledge of history may so inspire his students that they will gain a genuine knowledge and appreciation in his courses.

Here, then, lies the primary problem with peer rating. Too often, ratings are based upon "hearsay" passed among the faculty in the lounges. The

reason is that there is little or no opportunity for observation. It is clear, therefore, that the peer rating alone also does not provide a completely sound basis on which to determine merit. It is but a second facet in the overall evaluation.

Student Rating. The student rating is a third source of evaluation. This type of rating, however, is not as widely used as the first two because there seems to be considerable faculty opposition to student ratings. Many faculty members, as well as administrators, feel that students are not capable of judging the effectiveness of their instructors. The students, it is generally said, will not have a real opportunity to apply for some time what they have learned in the class. Moreover, it is contended that students do not have the maturity to realize the value of the teaching or of the course. In their discussion of student ratings, Remmers and Gage (1955: 492-501) point out six primary arguments against student ratings. These include the lack of judgment ability, the fact that teaching is not done at the pleasure of the student, snap judgments, the possibility of revenge for their own poor work or poor grade, the potential damage to faculty morale, and the misunderstanding which they may derive about their own "power" in controlling the teacher. In refuting these arguments, Remmers and Gage state that it is important to be aware of and to understand student attitudes; that student ratings are usually reliable and on the whole are not based upon grades earned. Student ratings conducted in the proper perspective serve to help the instructor identify weaknesses in his instruction so that he may work toward improvement.

In addition to refuting the arguments against student ratings, Remmers and Gage note that whether or not teachers like it, students inevitably will make their own evaluations of the instruction. Thus, by allowing the students to exercise the prerogative of evaluating the course, the instructor is

in a position to dispel rumor and campus gossip. The authors further explain that the time required to accomplish this task is relatively short, but the benefits which may result from the students' perceptions may prove to be highly valuable.

The use of student evaluations of teacher effectiveness is also supported by Churchill. (1966) In her discussion on teacher evaluation she stated "what the student believes is going on in a class is frequently very different from what the instructor intends to happen or believes is happening." Futhermore, she feels that "knowledge of the students' perception can help contribute to better teaching."

Student ratings obviously are not proposed as the best method of assessing teaching effectiveness. Few would agree that promotion or salary increases should be based only on a student opinion poll regardless of how responsible and mature the students might appear to be. Nevertheless, these ratings do provide another source to obtain an estimate of teacher effectiveness which may be utilized in conjunction with other sources to derive a more accurate picture of total competence or ability.

Self Rating. A fourth method of assessing effectiveness is that of self-rating. This concept of evaluation is frequently used as a means for instructional improvement whether a merit plan is in effect or not. Self-rating techniques provide the means for each individual to take stock of his work and to assess himself as objectively as possible.

It has been suggested frequently that in evaluating oneself, a person will ordinarily rate himself lower than his ability would indicate. For example, in the teaching profession a superordinate's appraisal of a faculty member is often higher than the individual's appraisal of himself. Whether it is a result of a lack of self-confidence or a sense of

humility which compels the person to underrate his own ability is a moot point; the fact that the self-rating score will generally be lower than the rating of a supervisor is the point to be noted.

On the other hand, as self-rating scales become more extensively used, faculty members should become more familiar with self-rating methods. Thus, the accuracy of self-ratings will improve and the tendency to underrate one's own performance should decline. In fact, it is quite possible that as faculty members begin to realize the real significance and the potential influence of their own self-rating upon salary increases or promotions, the rating scores will rise considerably. The scores in at least three situations with which the writer is familiar where self-rating had only an indirect effect upon promotions were, on the average, higher than the administrator - superior's ratings. Inevitably, some self-ratings will be too low, while others will be proportionately too high. Self-confidence and self-esteem are factors in this type of evaluation which cannot be controlled. From this, it is clear that self-rating cannot serve as the sole basis for determining merit increases in salary nor for promotions.

Combined Approach. In reviewing the advantages and disadvantages of each of the four rating groups discussed, it becomes evident that no one of them can adequately assess an individual's merit. While it is generally believed that administrative and peer ratings are normally conflicting, there is evidence also to indicate that peer ratings and administrator ratings do have a substantial correlation. (AASA, 1961: 32; Slocum, 1965) Still none of the rating scales by themselves can provide a complete picture of the individual or of his effectiveness in the institution. Combined, however, these estimates can provide the administrator with a sound basis for making decisions regarding merit increases and promotions.

A multiple rating approach would obviously be somewhat cumbersome and undoubtedly time-consuming. Yet, the extra time involved is a small price to pay for the dividends it can yield. In providing for different points of view, the accuracy of judgment would be increased in assigning merit salary increases and merit promotions to the most deserving individuals. Because of the increased accuracy and the provision for participation by several different groups, weaknesses in the evaluation program will be more clearly identified and better remedied. The resulting product will be a better faculty and an improved instructional program.

## Merit Programs--Success or Failure

The attempt to implement merit programs for pay increases and promotions is not new. Many have been tried and abandoned throughout the years, while others have met with success. A cursory review of only a few programs would provide the reader with a reasonably clear understanding of just why some programs succeeded while others had disastrous results.

Of prime importance in the success of any institution's program is the role of the instructional leader--the college president or the dean. Faculty members look to him to provide direction and guidance for progress and improvement of the academic program. He must be a strong administrator and dedicated to the profession, while at the same time he must be understanding and patient. He must cultivate change, not dictate; and he must believe in what he is doing, not merely jump on the bandwagon without reason or forethought. Above all, he must maintain the confidence of his faculty and solicit their advice and assistance whenever possible. Thus, the

overall success of any educational program, merit or otherwise, is directly dependent upon the qualities of the leadership in the institution.

A second factor which will have a direct bearing upon the success or failure of the program is the involvement of the faculty in its development. If faculty members are consulted, and if they share the responsibility for change, their efforts will most assuredly carry over into the implementation of a new program, for it reflects their thinking and their work. The administrator who fails to confer with his faculty and use their abilities fails to understand human relations and the nature of the group process.

The third point to be discussed here is that of continuous communication. Keeping lines of communication open prevents misunderstanding by keeping faculty members informed. It also precludes widespread circulation of rumors and hearsay. The successful operation of any institution requires that the people involved know what is going on about them. Open lines of communication promote morale by reassuring faculty members of their importance to the organization. This, in turn, provides the individual with a sense of worth and a sense of responsibility for the successful operation of the program.

In several instances where merit programs were put into practice effectively (Thorne, Alexander, Cushman, Bragg, Gores, and Bushong; 1957: 143-176) lines of communication were an essential element in the programs. In each case the faculty as a group or through representatives shared the responsibility for devising the merit plan. Moreover, the flow of information permitted faculty members to know what they could expect from the plan and what was expected of them. It was noted that in the majority of cases evaluation was not based solely upon the administrator's rating, but included ratings from two or three sources. All of the plans made provision

for the rated individual to have a part in his own evaluation. Finally, in each of the systems where merit plans met with success, it was obvious that the administrator was working toward the improvement of his faculty and soliciting their help in devising the merit program.

Merit plans were also abandoned in a number of cases. (NEA, 1957: 186-191) The reason for abandonment of some plans was the adoption of a single salary schedule; most of these, however, came about in the 1930's when, understandably, money was not available for extra pay. Other causes for failure included such reasons as subjective evaluation, misunderstanding within the faculty, and arbitrary limitation on the number of members who could receive merit raises, poor administrative judgments, partiality on the parts of the raters, and lack of available funds. The last reason identified is, nevertheless, not really a criticism of merit pay; rather it is merely a statement of an existing condition which would preclude any form of merit plan. Admittedly, effective merit pay-promotion plans can be superimposed only upon a sound basic salary schedule.

Ultimately it seems evident that the success of a merit plan, the same as the success of any educational plan, is dependent upon the combined efforts of the administration and faculty. Only through the harmonious involvement of all concerned can the task be achieved and the goal be reached.

## A Comparison of the Contents of Rating Scales

As stated earlier, no standard criteria for rating teacher effectiveness are accepted at this time. It is for this reason, then, that rating scales may be best determined or developed within the institution itself. Obviously a better job will be

done in establishing a program where guidelines are available to follow. This study has attempted to point out some guidelines for developing merit plans through the review of recent research and literature concerning the techniques used in merit rating and evaluation programs. In addition to the review, an investigation was made regarding the evaluative criteria used in twenty-one rating scales. These included seven "administrative/supervisory" rating scales, four "peer" rating scales, five "self" rating scales, and five "student" rating scales. Of the total, four "administrative," two "peer," three "self," and two "student" scales were derived from junior and community colleges; the remainder were included in reports and current literature. The analysis concerned the construction of a master list of evaluative items, the rank importance of the items indicated in each of the four types of scales, the comparison of the differences in the rank order among the groups, and the comparison of the combined ranks with those found in other studies.

Initially, each of the scales was tabulated item by item and then cross-matched to derive a single list of items representing the twenty-one scales. When combined, there were sixty-three separate entries.* (Appendix) Further examination revealed that these items could be subsumed under ten major categories. The ten categories were broad in nature and were structured in such a way that each could be superimposed over one or more of the more precise statements. Included in the ten major headings were: "classroom teaching," "personal attributes," "professional growth," "faculty-student relations," "community service," "service to the institution," "length of service," "research," "publication," and "competing offers."

---

* Some of the items stated on the individual rating sheets which were lengthy were shortened to facilitate the construction of a master list of major categories.

The categories were then ranked in importance within each of the four types of rating scales. (Table 2) The ranking of the categories was accomplished by tabulating the frequency with which the items in each category appeared in the twenty-one scales. Although the data were insufficient for any of the more sophisticated statistical techniques, comparisons of the rankings of the categories among the four types of rating scales yielded some notable findings.

First, and probably most outstanding, the analysis showed that classroom teaching was considered by far the most important aspect in faculty rating by all four of the groups. In terms of the number of items relating to instruction or classroom teaching, 84 of the 222 entries, or 37.8 percent, of the total pertained to this area. Moreover, by this approach it was considered more than twice as important as the second ranked category by both the administrative scales and the self-ratings, and more than four times as important as the second area by the student group. The peer rating indicated only a slight difference in emphasis for the first category over the second.

The second ranked category according to the combined tabulations for the four groups referred to "personal attributes." This category, however, was weighted as second only by three of the groups: in the administrative ranking it placed third behind "professional growth." This category accounted for 38 of the 222 items, or 17.1 percent. It should be noted that the peer group placed greater weight upon personal attributes in rating faculty members than any other group.

"Professional growth activities" was the third ranked category in the analysis. It held the second rank in the administrative rating scales, and tied for third place in the peer and self-ratings. The student ratings did not place any weight at all

TABLE 2
RANK IMPORTANCE OF EVALUATIVE CATEGORIES BY FOUR TYPES OF RATING SCALES

| Categories | Combined Totals | | | Administrative | | Peer | | Student | | Self | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rank | Number of Items | % of total | Rank | Number of Items | Rank | Number of Items | Rank | Number of Items | Rank | Number of Items |
| Classroom Teaching | 1 | 84 | 37.8 | 1 | 26 | 1 | 15 | 1 | 26 | 1 | 17 |
| Personal Attributes | 2 | 38 | 17.1 | 3 | 11 | 2 | 14 | 2 | 6 | 2 | 7 |
| Professional Growth | 3 | 28 | 12.6 | 2 | 15 | 3.5 | 9 | - | 0 | 3.5 | 4 |
| Faculty-Student Relations | 4 | 20 | 9.0 | 5 | 7 | 6 | 5 | 3 | 4 | 3.5 | 4 |
| Community Service | 5 | 17 | 7.7 | 7 | 5 | 3.5 | 9 | - | 0 | 6 | 3 |
| Service to the Institution | 6 | 14 | 6.3 | 5 | 7 | 5 | 6 | - | 0 | 8.5 | 1 |
| Length of Service | 7 | 8 | 3.6 | 5 | 7 | - | 0 | - | 0 | 8.5 | 1 |
| Research | 8 | 6 | 2.7 | 8.5 | 3 | - | 0 | - | 0 | 6 | 3 |
| Publication | 9 | 6 | 2.7 | 8.5 | 3 | - | 0 | - | 0 | 6 | 3 |
| Competing Offers | 10 | 1 | 0.5 | 10 | 1 | - | 0 | - | 0 | - | 0 |
| TOTALS | | 222 | 100.0 | | 85 | | 58 | | 36 | | 43 |

on this area as a means for evaluation. Totally, 28 items, or 12.6 percent were devoted to this area.

Ranking fourth among the ten evaluative categories was "student-faculty relations." In the four groups this classification was ranked fourth by the administrative group along with "service to the institution" and "length of service," sixth in the peer scales, third in the student group, and third in the self-rating scales along with "professional growth." In all, this category included 20 items, or 9.0 percent of the total.

The remaining six categories were ranked as follows in accordance with the combined frequencies of the items as they were specified in the four types of rating scales: (5) "community service" which accounted for 17 items or 7.7 percent of the total; (6) "service to the institution," to which 14 items were attributed or 6.3 percent; (7) "length of service" referred to in 8 items or 3.6 percent; (8) "research" and "publication" (tie), each of which accounted for 6 of the items or 2.7 percent; and (10) "competing offers" which was mentioned only once or in less than 0.5 percent of the total.

Regarding the ranking of the last six categories, there are several notations which seem to have special meaning. The student rating scales, for example were concerned with only three of the ten general categories: classroom teaching, personal characteristics, and faculty-student relations. None of these was among the latter six in the combined totals. Greater weight was placed instead on areas which would have the most bearing upon the students' own well-being. The peer group, too, indicated more of a concern with the items which seem to affect most directly the faculty and the institution than with such things as research and publication which would provide only an indirect benefit to the institution. The peers seemed to concentrate their

evaluation upon instruction, personal attributes, professional growth, and community service. Finally, it was particularly noticeable that only the administrative evaluations placed any weight upon competing offers, and even then it was mentioned only once.

## Implications

The significance of the foregoing analysis lies not in the production of a standard rating-scale checklist, nor in the establishment of universal criteria to be used in evaluating faculty members. The real benefit of the study accrues from the understanding that no absolute measurement of effectiveness exists. There are some guidelines, nevertheless, which can definitely facilitate the development of a sound rating device, and which can suggest some techniques for successfully implementing it.

The findings in this study indicate a high degree of agreement with those cited earlier by Gustad, and Astin and Lee. In computing the concordance, the relationship among the three rankings (Kendall's $W$) was found to be better than .91. (Table 3) The average correlation derived by formula from the correlations among the three surveys was computed to be .87.

The indications from all of these rankings call attention to the consistent pattern of beliefs among junior college administrators and faculty members (and to a degree among the students) about what is important in assessing effectiveness of college instructors. The identification of these factors and the proportionate importance they should have, consequently provide a sound basis around which an acceptable and effective rating scale can be built.

TABLE 3

COMPARISON OF RANK IMPORTANCE OF EVALUATIVE CATEGORIES

| CATEGORY | RANK | | |
|---|---|---|---|
| | Gustad | Astin - Lee | IHE* |
| 1. Classroom Teaching | 1 | 1 | 1 |
| 2. Personal Attributes | 2 | 2 | 2 |
| 3. Student Advising | 3 | 4 | 4 |
| 4. Committee Work | 4 | 5 | 6 |
| 5. Length of Service in Rank | 5 | 3 | 7 |
| 6. Professional Society Activity | 6 | 6 | 3 |
| 7. Public Service | 7 | 7 | 5 |
| 8. Publication | 8 | 8 | 8.5 |
| 9. Research Activity | 9 | 10 | 8.5 |
| 10. Competing Offers | 10 | 9 | 10 |

* The ranking of the Institute of Higher Education was adjusted to the terminology used in describing the factors listed by Gustad and Astin and Lee.

$W = .91$

$r_s = .87$ av.

Another finding in this study is that all groups fundamentally agree on what qulaities are essential for faculty effectiveness. That is, each tends to be concerned with production in the classroom, behavior traits, faculty-student relations, and professional growth. It is logical, then, that if these interests are common to each group a more accurate and adequate evaluation can be made by the use of two or three ratings from different sources. While at present there is no flawless means of evaluation, no universal panacea, the probabilities of accuracy in judgment can be improved greatly by receiving several independent ratings which would provide different perspectives for the total evaluation.

## Summary and Conclusions

The purposes of this study were threefold. The first purpose was to review research and literature pertaining to the concept of merit ratings for salary increases and promotions. The second purpose was to analyze several rating scales and evaluation procedures which had been used recently or are currently in use. The third and most important purpose of the study was to make suggestions and recommendations for developing and implementing merit plans.

At the beginning of this report it was indicated that evaluation is a recurring element in our society. In every field of endeavor evaluation takes place in one form or another. Salesmen are rated by the volume of sales; industrial employees are rated in accordance with units produced; professional personnel are rated with regard to the successes they have achieved in their fields; officers in the military are rated on their leadership abilities. Rating in one form or another is prevalent also in higher education and has been for a number of years.

Inasmuch as differences do exist in teaching or in faculty effectiveness it seems logical that a superior effort should be rewarded. The problem, however, has revolved around the means to identify that superior effort or that outstanding individual.

Throughout this paper it has been indicated that there is no clear-cut set of rules or standard pattern which may be used for any institution. Because of the complexity in measuring effectiveness and the differences in the roles played by the variety of institutions of higher education, criteria development has posed a difficult problem. The findings of this study indicate, nevertheless, that according to the type of institution there does appear to be agreement regarding what factors are important in judging effectiveness. In addition, these identified factors have been ranked in importance for the various types of institutions. They are not to be assumed to be precise criteria, but they can serve as guidelines or as the foundation around which the criteria may be developed by the institutions themselves. That is, the more explicit elements of merit rating plans should be developed in accordance with the general objectives of each individual institution.

The internal development of merit plans will allow for greater understanding within the faculty regarding what criteria are to be used in measuring effectiveness, and will provide for better acceptance of decisions based on such ratings. Furthermore, when faculty participation in the construction of the plan is provided, the responsibility for its success is shared both by the faculty and the administration.

The investigation of rating scales indicated that both rating checklists and descriptive phrases should be included in the overall evaluation. Neither type alone seems to be adequate. Numerical ratings do not always supply enough information while descriptive phrases are often too lengthy and too easily misinterpreted. Combined, however, they can

present a reasonably accurate picture of an instructor's effectiveness.

With regard to the question of who shall judge, research has failed to show where any one group, administrators, (including superordinates) peers, students, or faculty members themselves, holds a unique ability to evaluate faculty effectiveness. Each can contribute to the overall evaluation by providing a different point of view. The findings also indicate considerable agreement among the different groups (administrators, peers, students, and self) as to what is considered important in faculty ratings. While this information may have little significance on the surface, it implies that ratings made by the different sources have a relationship in that they attempt to evaluate similar elements. This means that more complete information can be obtained through the use of several independent sources. Accordingly, the accuracy of any decision based on merit should be improved considerably through the use of several ratings from different sources. Although the time needed to accomplish the task of evaluating faculty members may be increased, there is little doubt that the results yielded from the multi-rating approach will make it worthwhile.

In conclusion, evaluation is inevitable. Whether it is constructive or destructive, whether it is formal or informal, whether it is planned or unplanned, and whether it is equitable or unjust will depend upon the effort spent in building the plan. The success of any instructional program will be proportional to the support and the cooperation it is given by the faculty. This is similarly true in developing a merit plan. Time and again it has been illustrated that the involvement of the faculty in developing and carrying out new programs encourages both their support and cooperation. By developing the merit plan within the institution, using research and literature as guidelines only, the resulting evaluation program should provide the most comprehensible, the most applicable, and the most accurate measuring instrument available for that institution. Moreover,

it should provide a sound and acceptable basis for making decisions concerning merit salary increases and merit promotions. A merit plan can succeed only as a cooperative endeavor of the administration and the faculty. The resulting product will be a better faculty and an improved instructional program.

## SELECTED BIBLIOGRAPHY


1. Alexander, R. Virginia, "Teacher Evaluation Program: Ladue Public Schools", Journal of Teacher Education, Vol. III, No. 2, NEA, Washington, D.C., 1957, 148-153

2. Astin, Alexander W.; Lee, Calvin B.T.; Kent, Laura, "The Evaluation of Teaching", The Educational Record, American Council on Education, Washington, D.C., 1966, 361-406

3. Barr, A.S., "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations", Journal of Experimental Education, Vol. XVI, Dembar Publications, Inc., Madison, Wisconsin, June 1948, 203-283

4. Burton, William H., Brueckner, Leo J., Supervision, A Social Process, Appleton-Century-Crofts, Inc., New York, 1955

5. Bushong, James W., "The Story Behind Grosse Pointe's Professional Growth Program", Journal of Teacher Education, Vol. III, No. 2, NEA, Washington, D.C., 1957, 170-175

6. Churchill, Ruth, "Student Ratings of Faculty", (Paper delivered on Panel on Evaluating Teacher Effectiveness, American Association of Junior Colleges), March 2, 1966, 1-4

7. Cushman, Jack, "The Glencoe Career-Teacher Plan", Journal of Teacher Education, Vol. III, No.2, NEA, Washington, D.C., 1957, 154-158

8. Ellena, William J. et al (Ed.), Question, Who's a Good Teacher?, American Association of School Administrators, NEA, Washington, D.C., 1961

9. Engleman, Finis, "Difficulties and Obstacles Inherent in Merit Ratings for Teachers", Journal of Teacher Education, Vol. III, No. 2, NEA, Washington, D.C., 1957, 136-139

10. Gores, Harold B., "Awards For Notable Service", Journal of Teacher Education, Vol, III, No. 2, NEA, Washington, D.C., 1957, 165-169

11. Gragg, W. L., "The Ithaca Merit Salary Program", Journal of Teacher Education, Vol. III, No. 2, NEA, Washington, D.C., 1957, 159-164

12. Graybeal, William S., "What the College Faculty Thinks", NEA Journal, Vol. 55, No. 4, Washington, D.C., April, 1966, 48-49

13. Gustad, John W., Policies and Practices in Faculty Evaluation, American Council on Education, Washington, D.C., 1961, (monograph)

14. Mitzel, Harold E., "Teacher Effectiveness", Encyclopedia of Educational Research, 3d ed., Macmillan & Co., 1960, 1481-1485

15. Remmers, H.H., Gage, N.L., Educational Measurement and Evaluation, Harper and Brothers, New York, 1955

16. Slocum, Clyde Wesley, Dissertation: "An Exploration of the Perceptual Relationship Between Peer and Principal Ranking of Teachers", University of Illinois, 1965

17. Stinnett, T.M. Ed., "Why Merit Salary Schedules Were Abandoned", Journal of Teacher Education, Vol. III, No. 2, NEA, Washington, D.C., 1957, 186-191

18. Thorne, Edmund H., "West Hartford's Career Salary Plan", Journal of Teacher Education, Vol. III, No. 2, NEA, Washington, D. C., 1957, 143-147

19. Wiles, Kimball, Supervision for Better Schools, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1955

20. Yamamoto, Kaoru, "Evaluating Teacher Effectiveness: A Review of Research", Journal of School Psychology, Vol, II, No. 1, Division of Special Education, Ohio Department of Education, Columbus, Ohio, 1963-64, 60-69

## APPENDIX
## ITEMS LISTED IN TWENTY-ONE RATING SCALES

1. Speaking voice
2. Mannerisms and pleonasms
3. Knowledge of subject
4. Clarity of presentation
5. Level of comprehension adjusted
6. Enthusiasm (self)
7. Enthusiasm engendered in class
8. Stimulates thinking
9. Encourages participation
10. Guides discussion
11. Digressions
12. Organization and preparation
13. Student responsibilities clearly set
14. Class time well spent
15. Use of analogies, illustrations, and examples
16. Handling of questions
17. General atmosphere created
18. Interested in students, i.e. patient, willing to help, etc.
19. Discipline problems - handling, absence of etc.
20. Courtesy
21. Consultative availability (students)
22. Integration of material (to other discipline areas)
23. Toleration of opposed opinion
24. Value of textual materials
25. Grading policies
26. Amount of outside work required
27. General effectiveness
28. Service to institution
29. Departmental service
30. Committee/Administrative work

---

* The twenty-one scales used here include seven "administrative/supervisory" scales, four "peer" scales, five "self" scales, and five "student" scales.

APPENDIX (Continued)

31. Extracurricular work
32. Academic achievement
34. Work on degrees in related fields
35. Professional growth
36. Professional activities
37. Professional ethics
38. Professional leadership
39. Professional status
40. Conventions
41. Recognition
42. Student/teacher relations
43. Humanitarian attitudes
44. Considerate of others
45. Relates well with community and colleagues
46. Promotion of institutional goals
47. Personal characteristics
48. Cooperativeness
49. Personal appearance, grooming
50. Easy to get along with
51. Personal outlook and attitudes toward profession
52. Friendliness
53. Community service
54. General teaching ability
55. Length of service to institution
56. Research
57. Publication
58. Length of career
59. Competing offers
60. Supervision of graduate study
61. Supervision of honors programs/students
62. Intelligence (as measured by standardized tests of mental ability and achievement)
63. Reflection of teaching ability in student changes

PUBLICATIONS
of the
Institute of Higher Education

UNIVERSITY ASSISTANCE TO DEVELOPING INSTITUTIONS
Galen N. Drewry and Durward Long. Price $1.00

THE FACULTY CRISIS: CUES FOR THE SMALL COLLEGE
Galen N. Drewry and Dean Baugher. Price 50¢

FACULTY PERCEPTIONS OF THE RESEARCH ENVIRONMENT
Cameron Fincher. Price 50¢

INSTITUTIONAL RESEARCH IN GEORGIA COLLEGES
Cameron Fincher. Price 50¢

COLLEGE ADMINISTRATION AND THE FACULTY HANDBOOK
Durward Long. Price $1.00

PROBABILISTIC VERSUS DETERMINISTIC MODELS IN COLLEGE ADMISSIONS
Cameron Fincher. Price 50¢

PLANNING IN HIGHER EDUCATION
Cameron Fincher. Price 50¢

ADMINISTRATIVE TEAM LEADERSHIP IN CONCEPT AND PRACTICE
Conwell A. Anderson, Editor. Price $1.00

THE COLLEGE NIGHT PROGRAM IN GEORGIA HIGH SCHOOLS
John J. Sullivan. Price 50¢

Additional copies of this report may be secured
from the Institute of Higher Education
Univerisity of Georgia, Athens, Georgia
for 50¢ each.